Приемы повышения точности геолокации телефонов на сети мобильного оператора

Каледин Артем @ билайн



О спикере

Каледин Артем



ВМК МС @ МГУ
1й выпуск MADE @ VK



> 5 лет в анализе данных

Делал проекты с гео в банках, телекомах, для такси/курьеров и маркетплейсов

Работал в логистическом стартапе и строил модельки для Икеи

2


Сколько базовых станций на фото?

План доклада

1 Какую задачу решаем, какие данные используем, что такое БС, где применяем решение

2 Базовые подходы по имеющимся данным и проблемы, с которыми столкнулись

3 Графы и гексагоны

4 Что использовали и следующие шаги



Какая ценность решения задачи внутри билайн



Нагрузка на сеть и планирование развития сети



«Точки притяжения» абонентов



B2B на обезличенных данных: открытие торговых точек и проходимость зон

Ссылка на открытое решение

geo-interface.com/

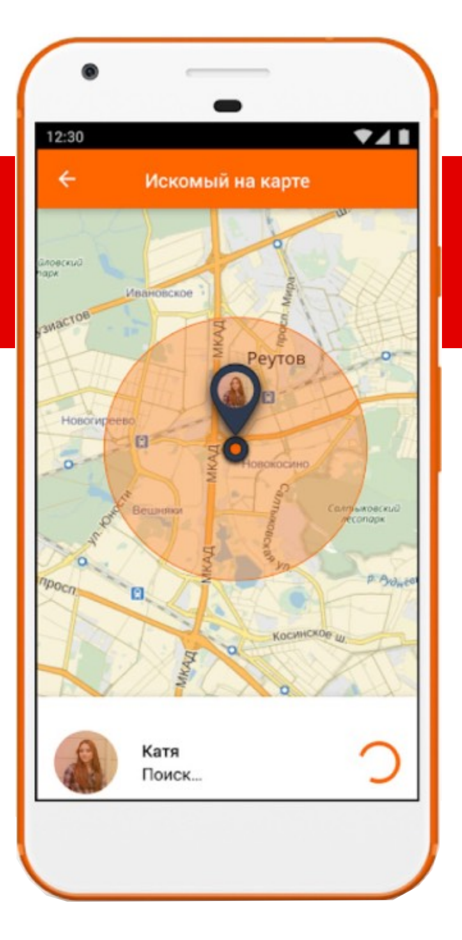




Социальная деятельность с ЛизаАлерт

- **билайн.поиск** – платформа на основе анализа больших данных, где приходит смс «свидетелям пропажи» человека
- **нейросеть Beeline AI** – ИИ отсматривает фото, полученные с дронов во время поисковой операции, и ищет на них пропавшего человека
- **пилот с Пятерочкой** в Москве для настройки сообщений на удалённых рекламных носителях о пропавших людях в зоне нахождения магазина

>100
людей





Мы работаем не с GPS, а со вспышками

И они есть везде!



Магазины

Заказы в конкретные пункты выдачи и посещения магазинов



Банки

Информация с POS-терминалов и входы в мобильные приложения



Такси/курьеры

Заказы, рестораны, места подачи и окончания поездок



____

Впишите свой бизнес и какие у вас есть гео-данные



Базовая станция

Это комплекс антенн, которые принимают и передают сигналы от телефона к телефону.

- Высота крепления — радиус покрытия
- Много пользователей — высокая пропускная способность



Какие БС бывают?

Фемтосоты







Макросоты





Какие есть данные

LAC

Код местности
(Location Area Code)

CELL

Код базовой станции
(сота)

lat, lon

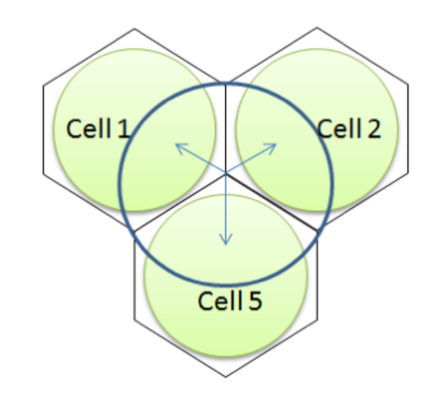
Координаты БС

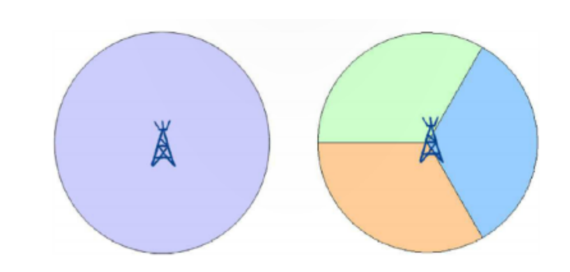
Зона свечения

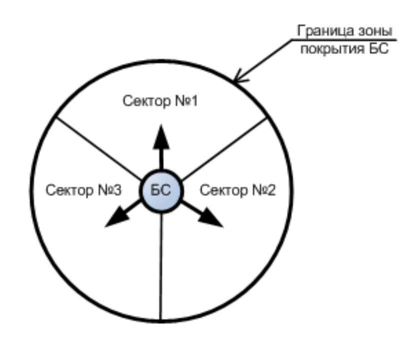
(сектор)

Для валидации

GPS-треки драйв-тестов









Гео = круто



Много данных



Частичная открытость

Пробки, OSM, индекс привлекательности мест, цены на недвижимость



Деперсонализированность и агрегационная способность

Нам не интересен конкретный клиент для открытия новой точки — нам интересно их количество



Триангуляция — вычислим ли точный адрес?

Триангуляция — разбиение геометрического объекта на симплексы

В случае плоскости — разбиение на треугольники

В сотовой связи — вершины заданы координатами базовых станций





Триангуляция — вычислим ли точный адрес?

Минусы:

- Перегруженность БС и перескоки
- Преломление сигнала и замер фактического расстояния
- Высокая погрешность. Не видим абонентов на подконтрольной территории





Какие есть данные

LAC

Код местности
(Location Area Code)

CELL

Код базовой станции
(сота)

lat, lon

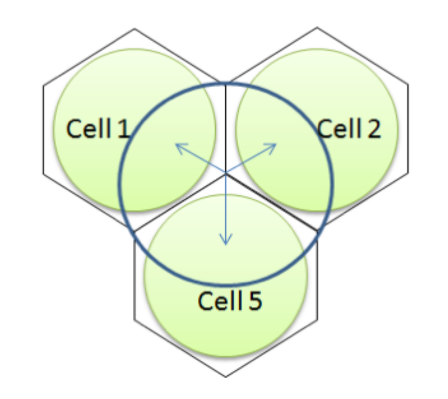
Координаты БС

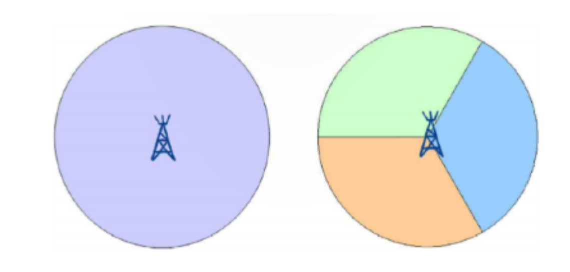
Зона свечения

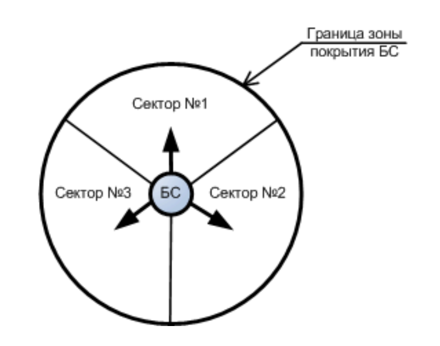
(сектор)

Для валидации

GPS-треки драйв-тестов









Как бы вы решали эту задачу?





1. Каждая координата БС – итоговая локация

БС

Число вспышек на БС

Переключения

GPS-трек

Примеры треков



Как считаем метрики?

1 Берем временное окно заданной длины

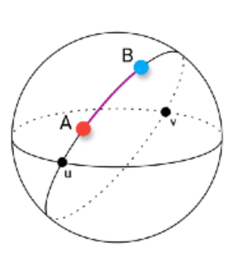
2 Агрегируем данные по ctn + ts window (обычно данные приходят каждую секунду => получаем истинное значение в окне)

3 Считаем медианные координаты для окна

4 Считаем расстояние по хаверсинусу

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\phi2 - \phi1}{2}\right) + \cos(\phi1)\cos(\phi2)\sin^2\left(\frac{\lambda2 - \lambda1}{2}\right)}\right)$$

Haversine

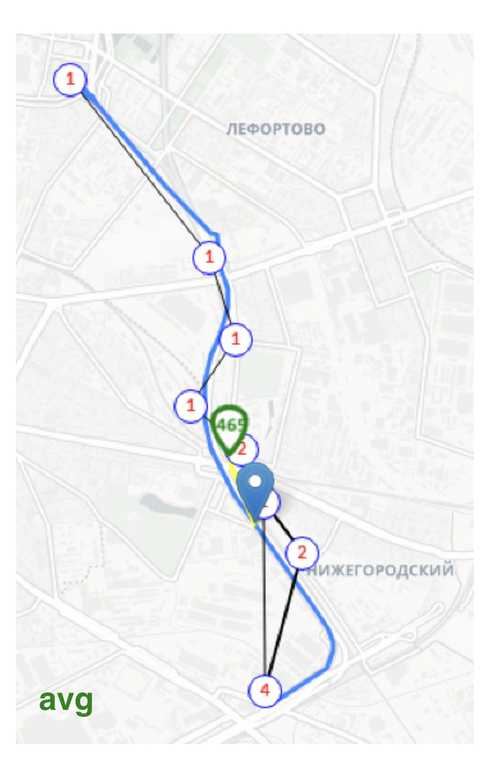


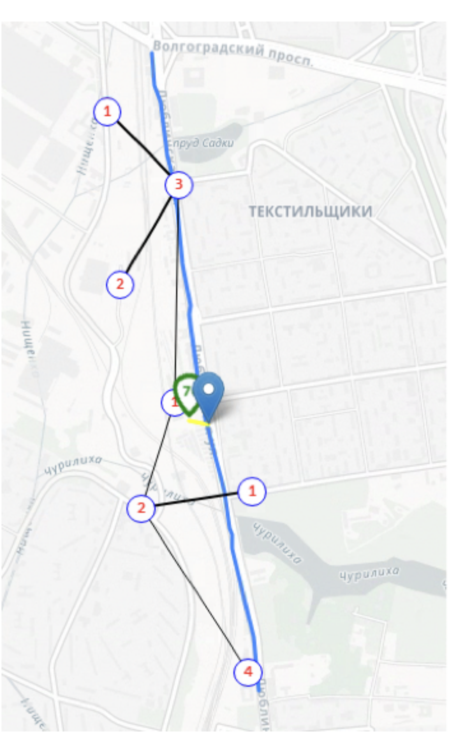
Хорошая геопозиция!

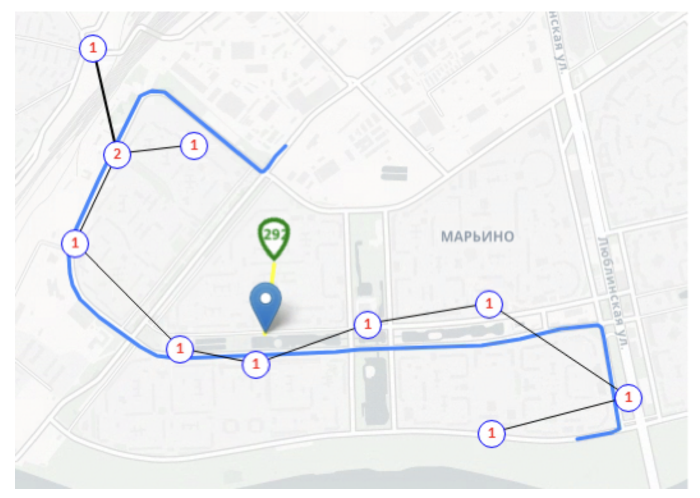
Плохая геопозиция

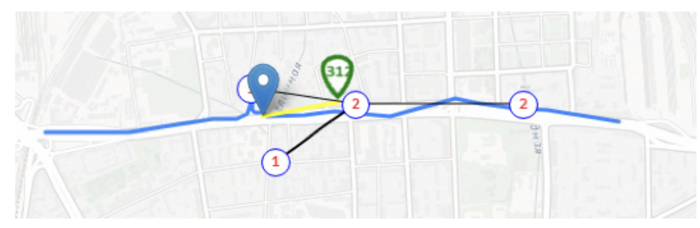
Агрегация по временным окнам:

2. Усреднение координат БС









Формула

$lat_{BS1} + lat_{BS2} + \dots lat_{BSN} / N$

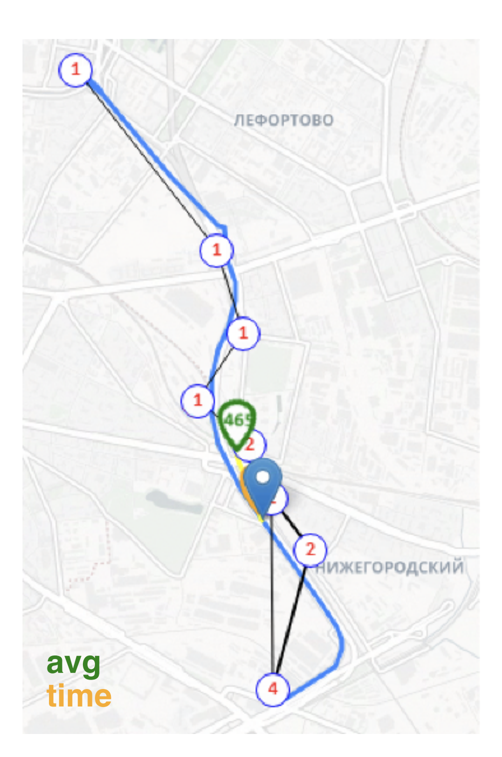
$lon_{BS1} + lon_{BS2} + \dots lon_{BSN} / N$

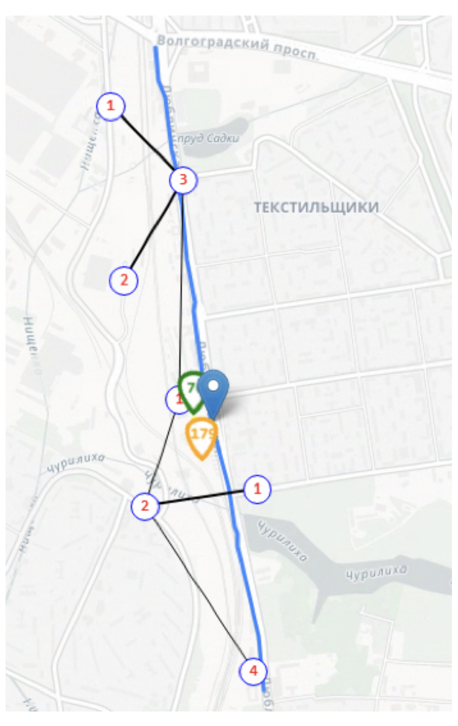
Примем значение метрики для этого случая за бейзлайн

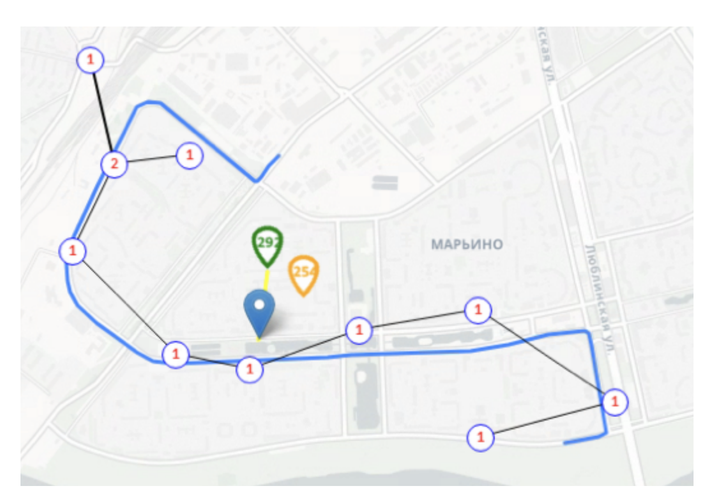


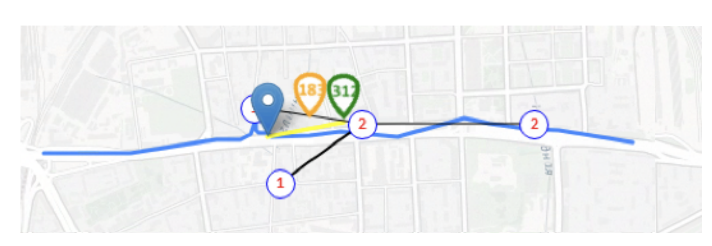
Агрегация по временным окнам:

3. Взвешенное по времени усреднение координат БС









Добавляем веса вида 1/t и ln 1/t

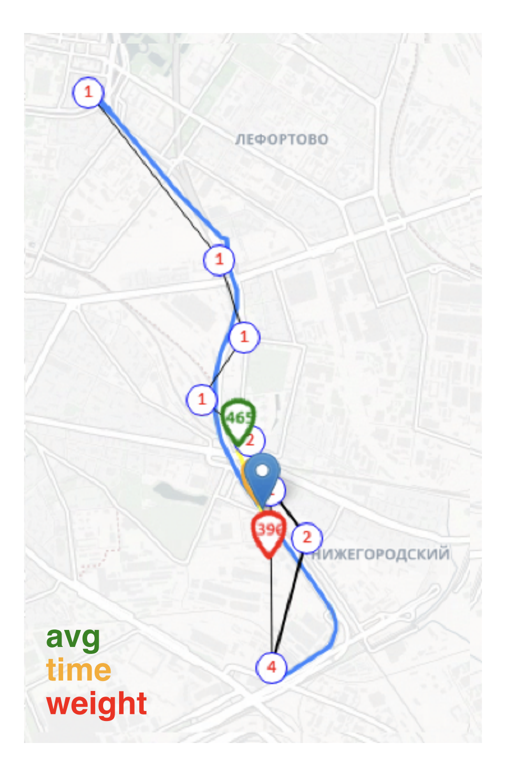
В 29% случаев метрика лучше бейзлайна

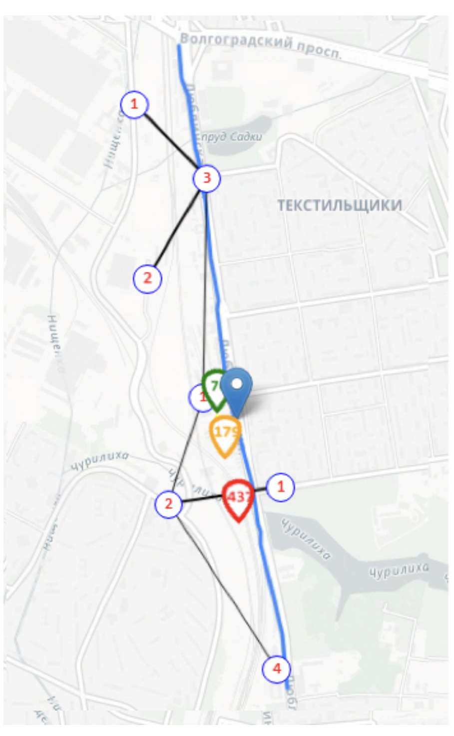
В 1.35 раз в среднем расстояние меньше

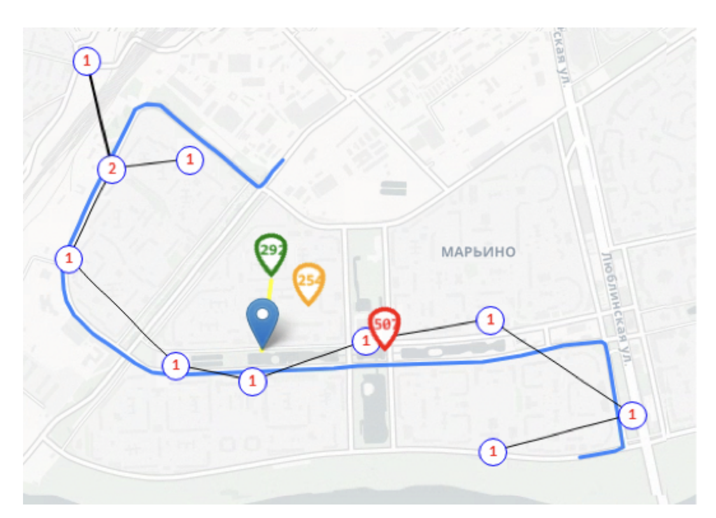


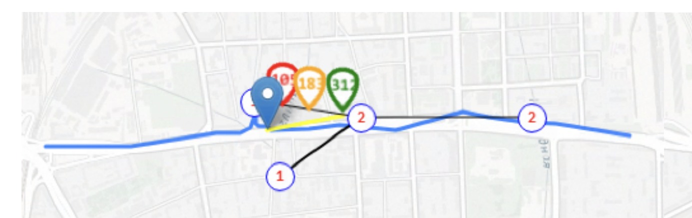
Агрегация по временным окнам:

4. Взвешенное усреднение координат БС









Добавляем веса вида 1/ cnt и ln 1/cnt

+ решаем линейной моделью

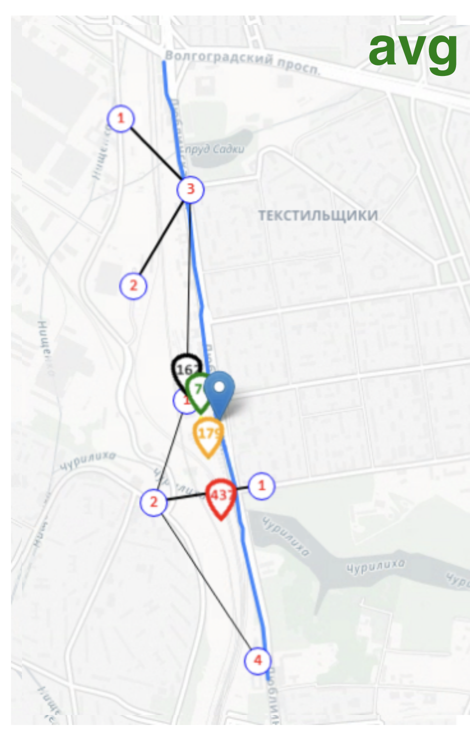
В 24% случаев
метрика лучше бейзлайна

В 1.9 раз
в среднем расстояние меньше

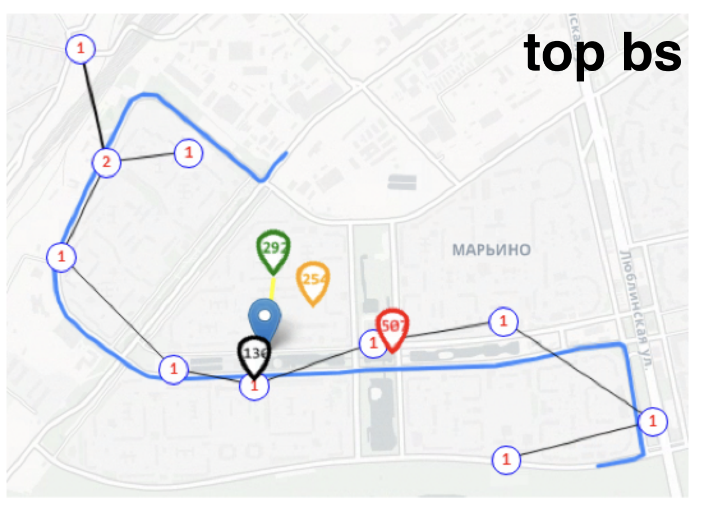


Агрегация по временным окнам:

5. Определение топ БС

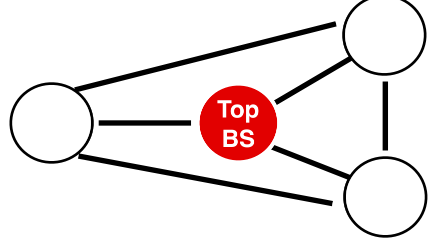


avg





Наименьшее расстояние до других БС



В 37% случаев

метрика лучше бейзлайна

В 1.8 раз

в среднем расстояние меньше



Сравнение методов

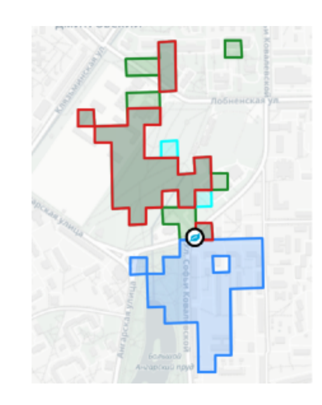
Название алгоритма	Усреднение	Взвешенное по времени	Взвешенное по счетчикам	Топ БС
Доля наилучших результатов	44 %	9 %	16 %	31 %

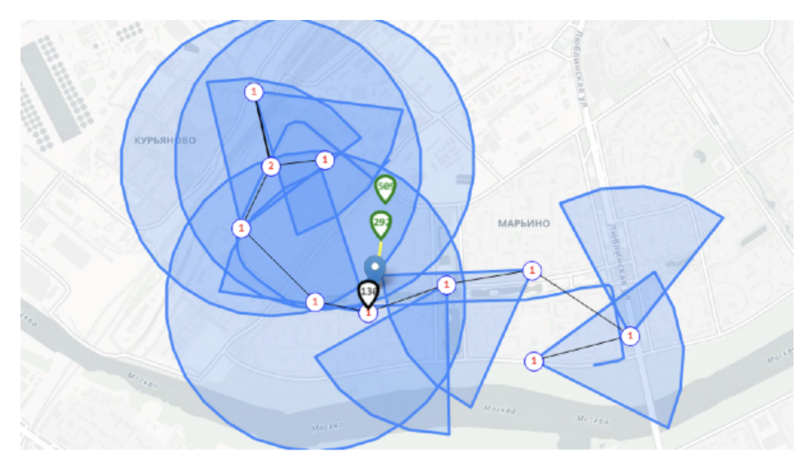


1. Центроид пары пересечений БС

- Попарно рассматриваем пересечения покрытий БС
- Выбираем наибольшее парное пересечение
- Считаем центроид образовавшейся геометрии


Пересечение двух БС


Пример 3 cell для 1 lac и их пересечение

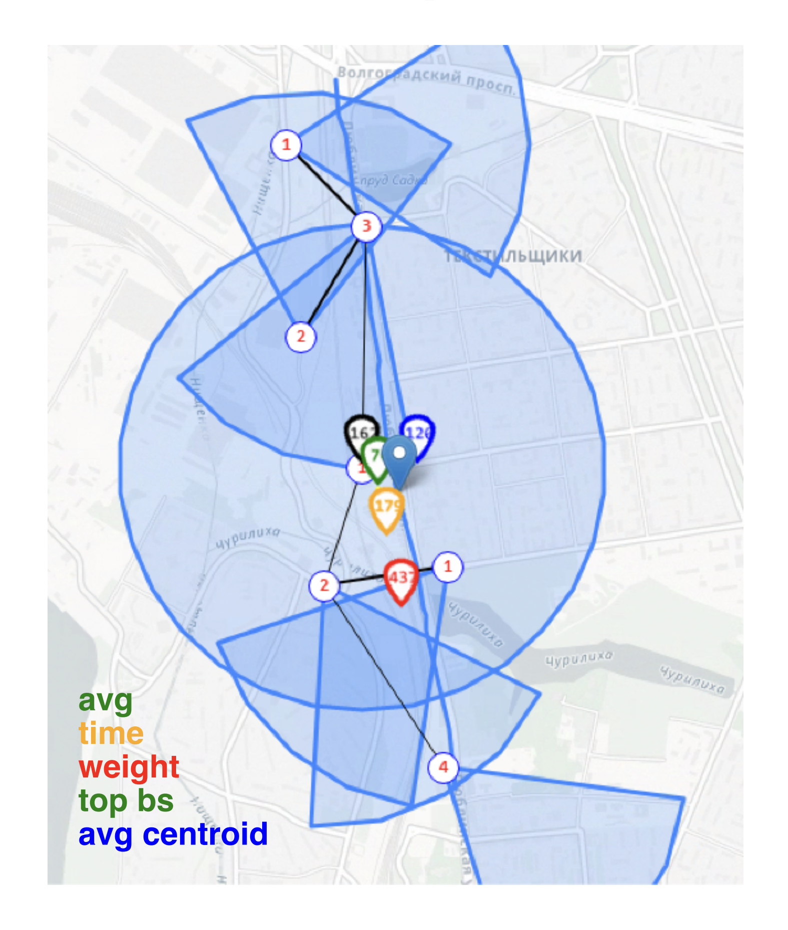


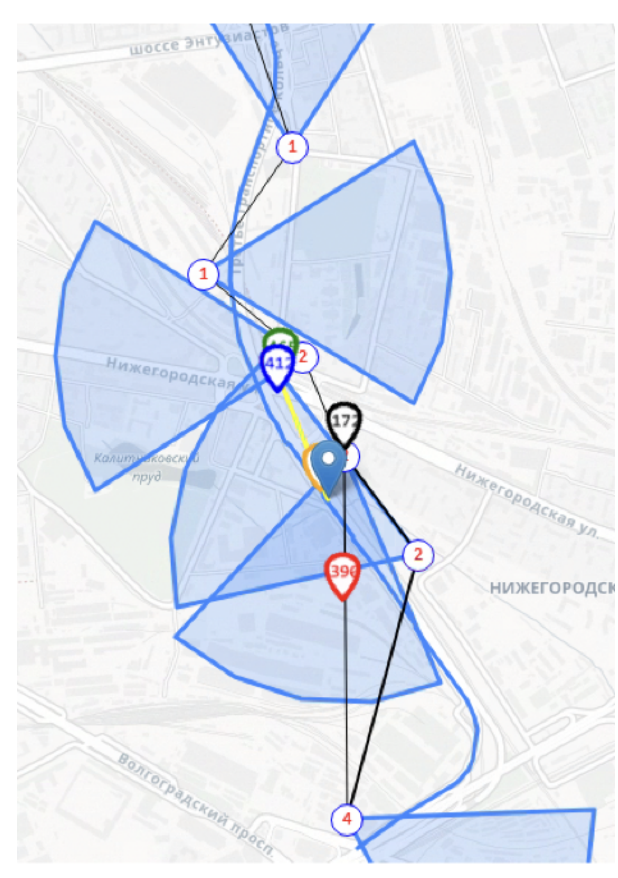
Пример теор. и факт. покрытий



Обработка свечений:

2. Усреднение центроидов геометрий





В 43% случаев

метрика лучше бейзлайна

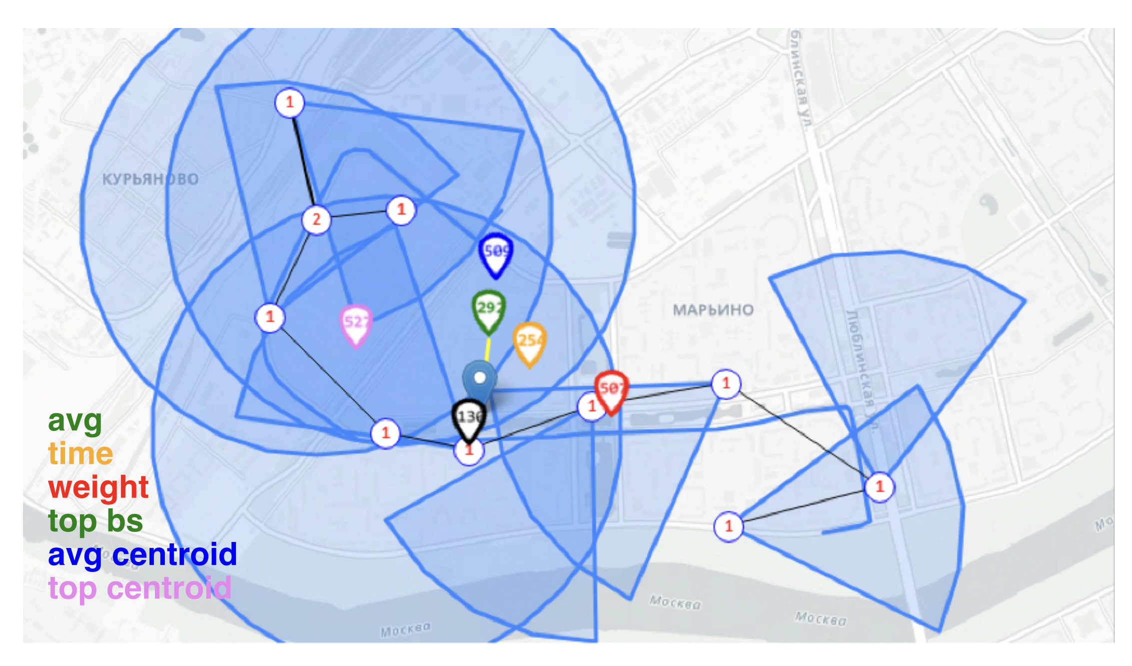
В 1.6 раз

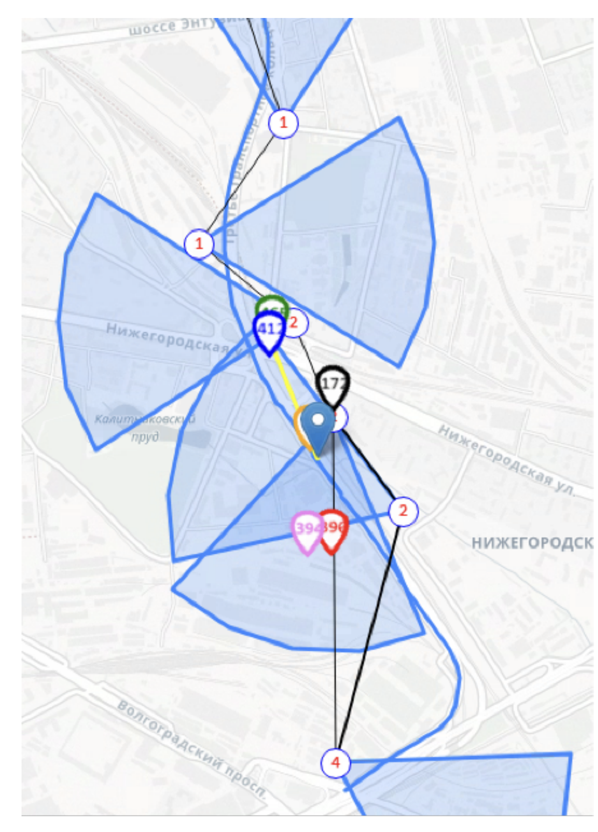
в среднем расстояние меньше



Обработка свечений:

3. Центроид геометрии топ БС





В 19% случаев
метрика лучше бейзлайна

В 1.8 раз
в среднем расстояние меньше



Сравнение методов

Название алгоритма	Усреднение	Взвешенное по времени	Взвешенное по счетчикам	Топ БС	Топ БС центроид	Усреднение центроидов	Линейные веса
Доля наилучших результатов	21 %	7 %	12 %	17 %	7 %	19 %	17 %



4. Проекция координат БС на граф

Обработка свечений базовых станций

- Знаем покрытие БС
- Берем дорожный граф в этой зоне
- Проецируем координаты БС на вершины графа





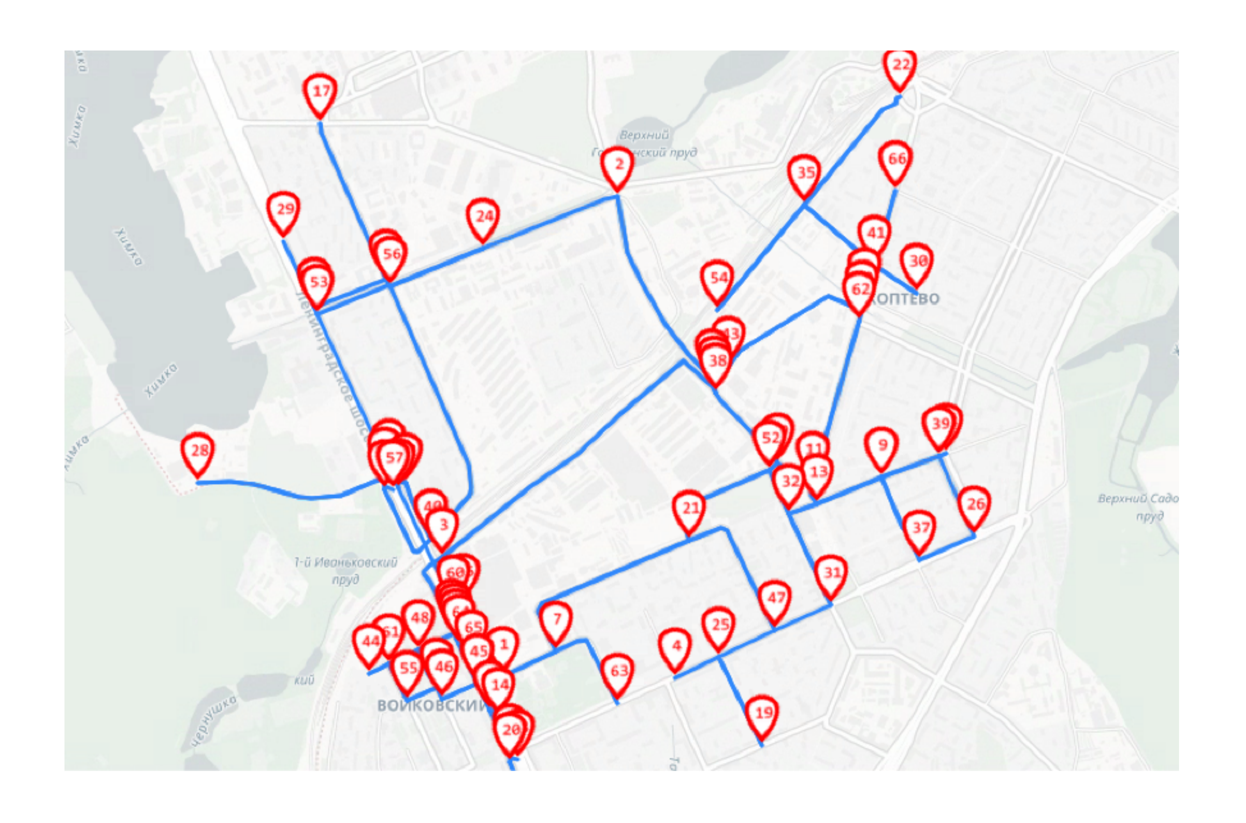
Делаем buffer, чтобы нашлись точки на графе



Построение маршрутов для графа дорог

Пытаемся восстановить по ребрам и вершинам дорожного графа маршрут через БС

Проецируем свечение БС или координату расположения на граф и строим кратчайший путь до следующих БС, тем самым создавая возможный трек абонента



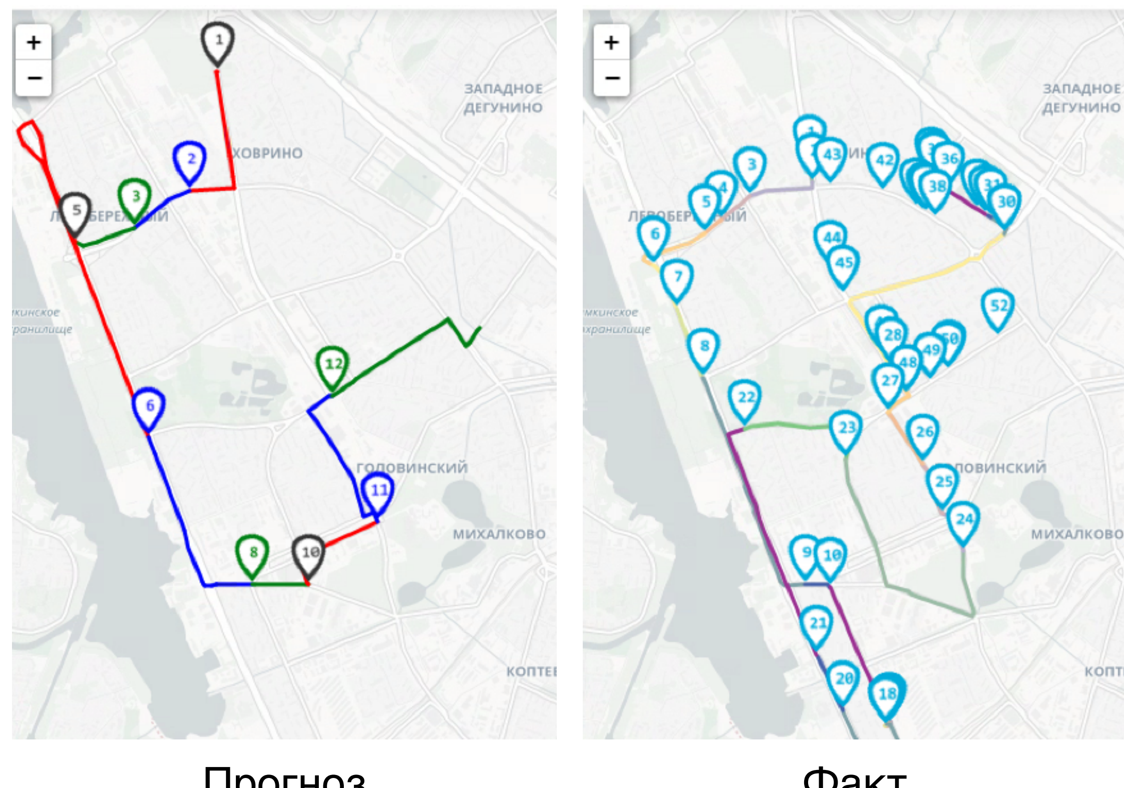


Построение маршрутов для графа дорог

Пытаемся восстановить по ребрам и вершинам дорожного графа маршрут через БС

Проецируем свечение БС или координату расположения на граф и строим кратчайший путь до следующих БС, тем самым создавая возможный трек абонента

В среднем работает на 15% лучше, чем прошлые методы, но вычислительно дорого




Прогноз Факт

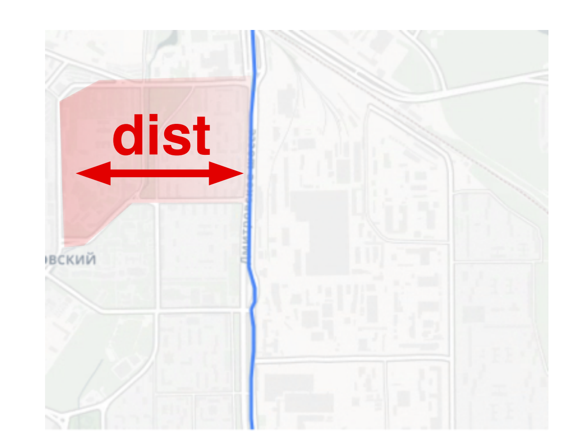


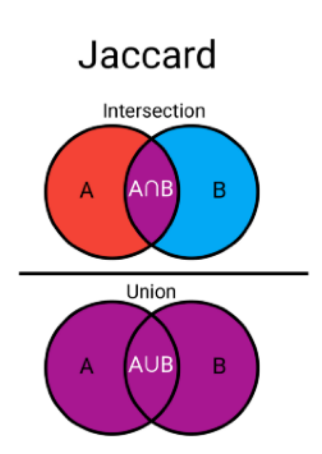
Как считаем метрики?

1 На отрезке имеем два трека — оригинальный и восстановленный

2 Считаем долю перекрытых областей

3 В зонах отличий считаем среднюю ошибку по хаверсинусу





Хорошая геопозиция!

Плохая геопозиция

HighLoad++ 2022


Сколько базовых станций на фото?

С какими проблемами сталкиваемся

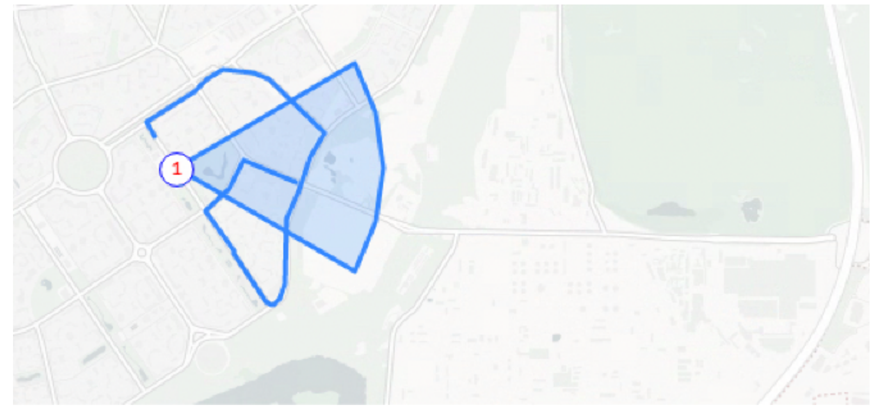
- Выбросы по GPS — линейные преобразования и фильтр Калмана



С какими проблемами сталкиваемся

- Выбросы по GPS — линейные преобразования и фильтр Калмана
- Сдвиги по времени



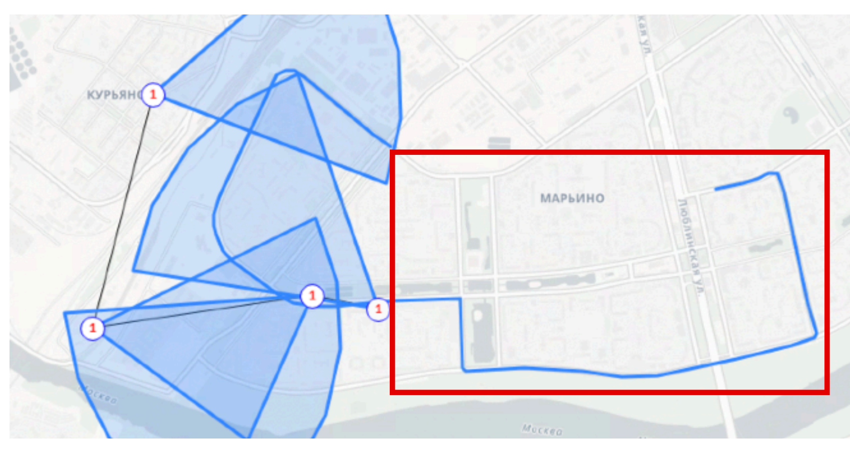


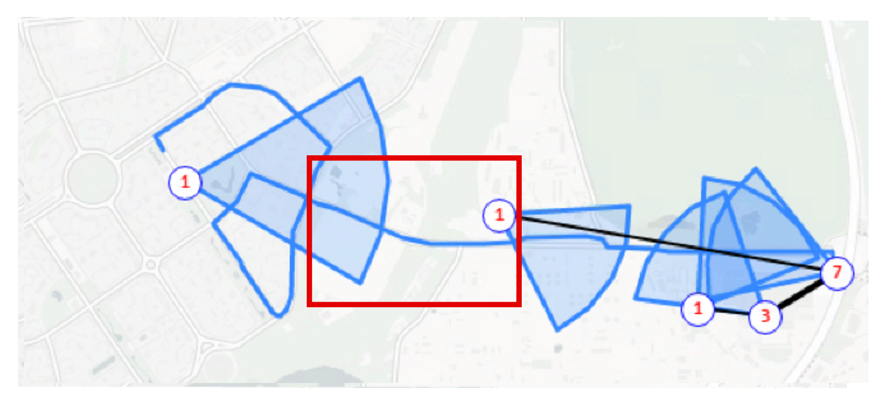


С какими проблемами сталкиваемся

- Выбросы по GPS — линейные преобразования и фильтр Калмана
- Сдвиги по времени







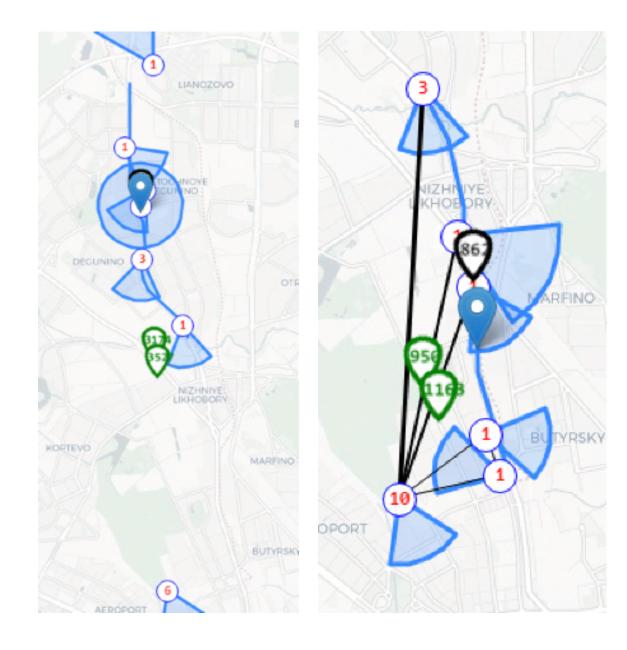
Запаздывание



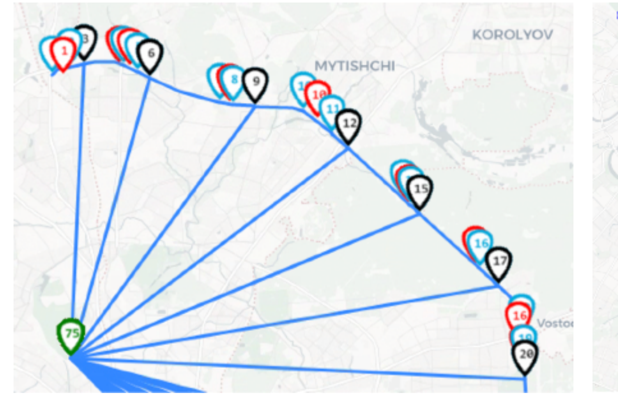
С какими проблемами сталкиваемся

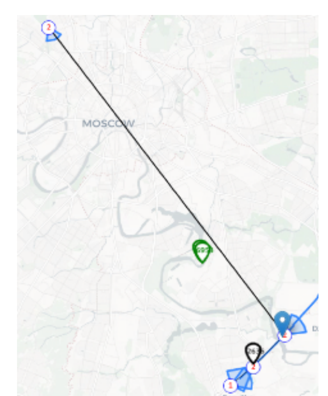
- Выбросы по GPS — линейные преобразования и фильтр Калмана
- Сдвиги по времени
- Выбросы

Неочевидные



Очевидные

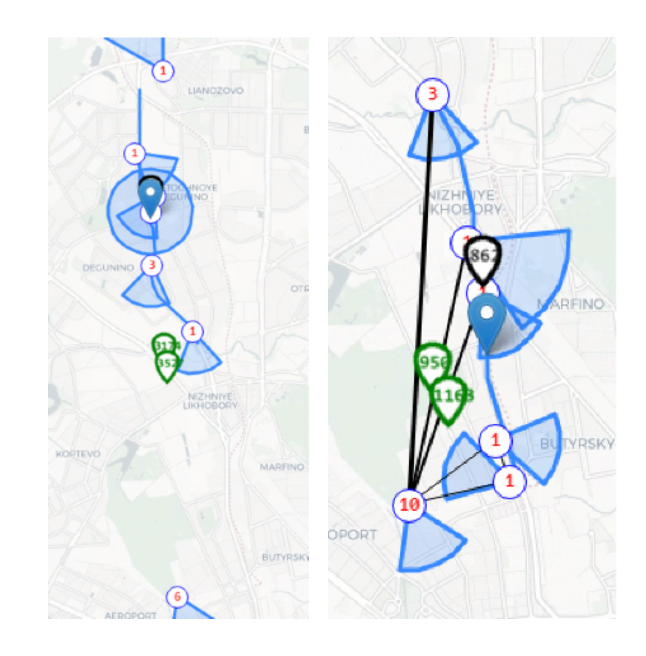




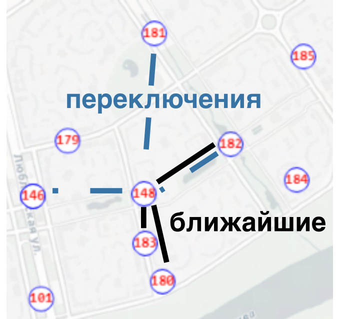


С какими проблемами сталкиваемся

- Выбросы по GPS — линейные преобразования и фильтр Калмана
- Сдвиги по времени
- Выбросы
  - Фактические расположения БС друг к другу и практические переключения БС



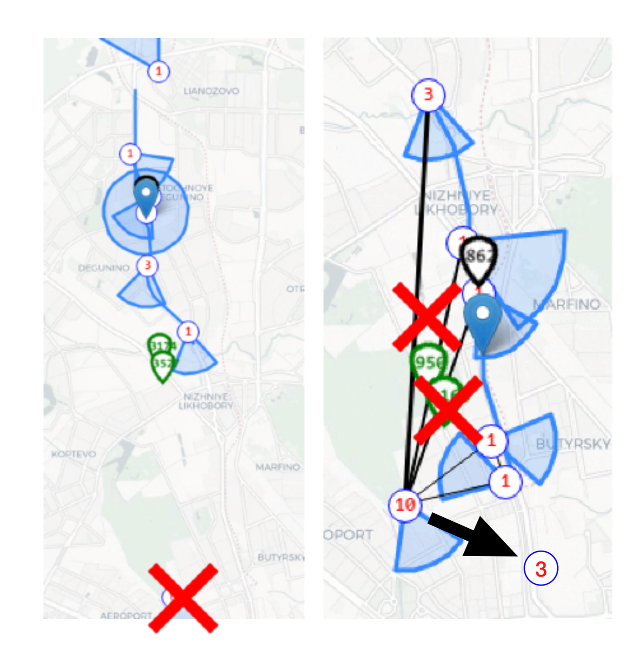




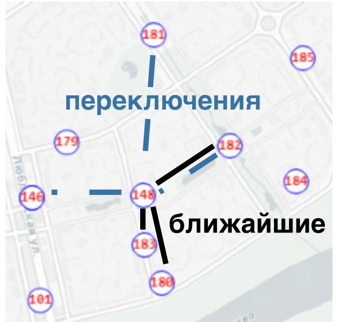


С какими проблемами сталкиваемся

- Выбросы по GPS — линейные преобразования и фильтр Калмана
- Сдвиги по времени
- Выбросы
  - Фактические расположения БС друг к другу и практические переключения БС



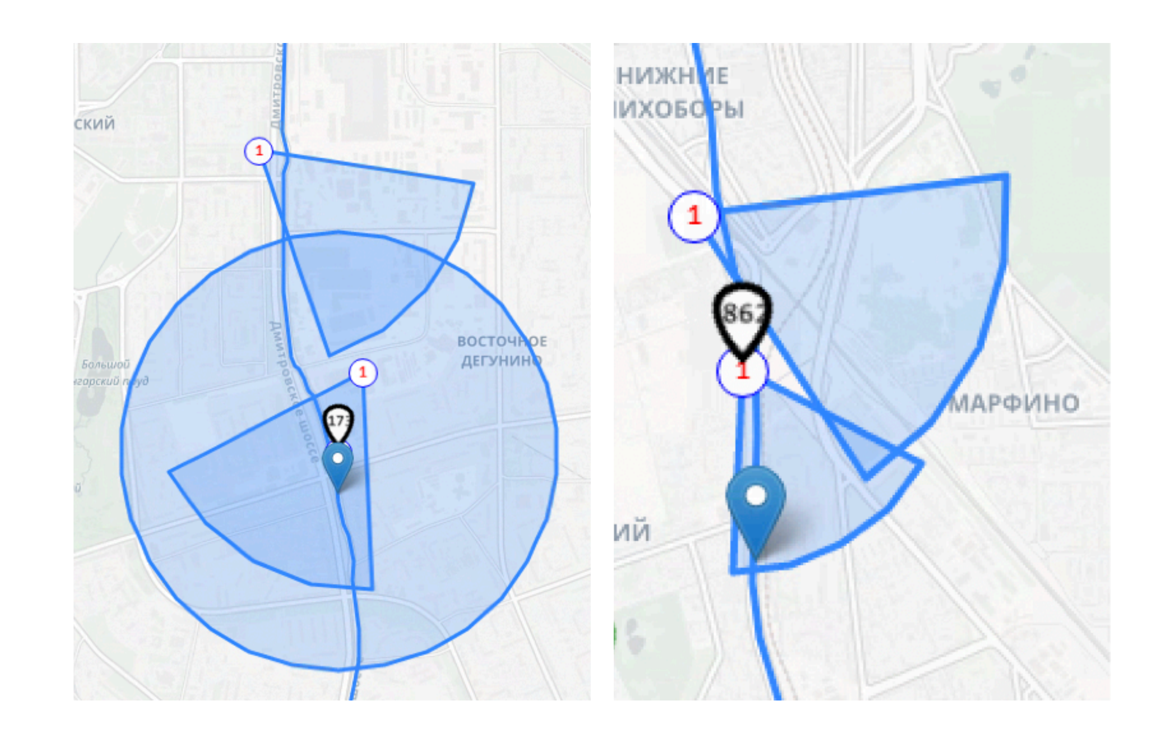






С какими проблемами сталкиваемся

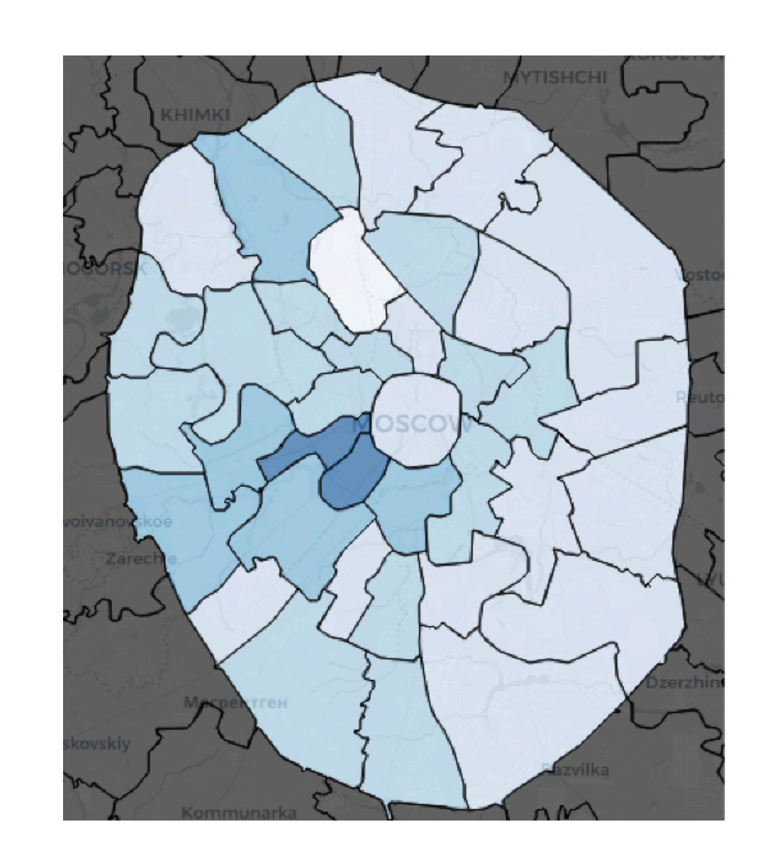
- Выбросы по GPS — линейные преобразования и фильтр Калмана
- Сдвиги по времени
- Выбросы
  - Фактические расположения БС друг к другу и практические переключения БС
  - Отсечение топ N БС по расстояниям





С какими проблемами сталкиваемся

- Выбросы по GPS — линейные преобразования и фильтр Калмана
- Сдвиги по времени
- Выбросы
  - Фактические расположения БС друг к другу и практические переключения БС
  - Отсечение топ N БС по расстояниям
  - Отсечение по медианному расстоянию. Анализ ошибок по районам и зашивание доли в район





С какими проблемами сталкиваемся

- Выбросы по GPS — линейные преобразования и фильтр Калмана
- Сдвиги по времени
- Выбросы
  - Фактические расположения БС друг к другу и практические переключения БС
  - Отсечение топ N БС по расстояниям
  - Отсечение по медианному расстоянию. Анализ ошибок по районам и зашивание доли в район
  - Пересечение проекций + изохроны




Сколько базовых станций на фото?

Точка притяжения

Место, в котором абонент проводит свое время (дом, работа, встречи с родственниками и друзьями)

Ее основные характеристики — **регулярность** и **продолжительность** нахождения в окрестности



Парадигма Stay/Move

Точка притяжения

Место, в котором абонент проводит свое время (дом, работа, встречи с родственниками и друзьями)

Ее основные характеристики — **регулярность** и **продолжительность** нахождения в окрестности

Stay	Move
Точки притяжения и разовые активности абонент находится в состоянии покоя и никуда не перемещается, но может переключаться между соседними БС	**Дороги, метро и т.д.** абонент находится в состоянии движения, обычно цель этого передвижения — добраться до другой точки притяжения



Как определить, на чем перемещается абонент?

- Оценка скорости перемещения по переключению между БС
- Сегментация БС — дороги, indoor, высота подвеса и обслуживаемая зона
- Выделение маршрутов общественного транспорта и сопоставление маршрутов с перемещением абонента



h3 index	Автомагистраль	Парк	Коммерческая зона	Школа	…
8a11818089a7fff	1	0	1	0	
8a11818090a7fff	0	1	1	0	
8a11818091a7fff	0	0	0	1	
8a11818092a7fff	0	0	0	0	



Что происходит с перемещениями?

Рассмотрим подробнее перемещение на авто

- При перемещениях мы можем предугадывать точку назначения и прогнозировать путь
- За счет справочника соседства (фактического и эмпирического) можем предсказывать набор следующих БС (де-факто марковские цепи)



Движение по магистрали

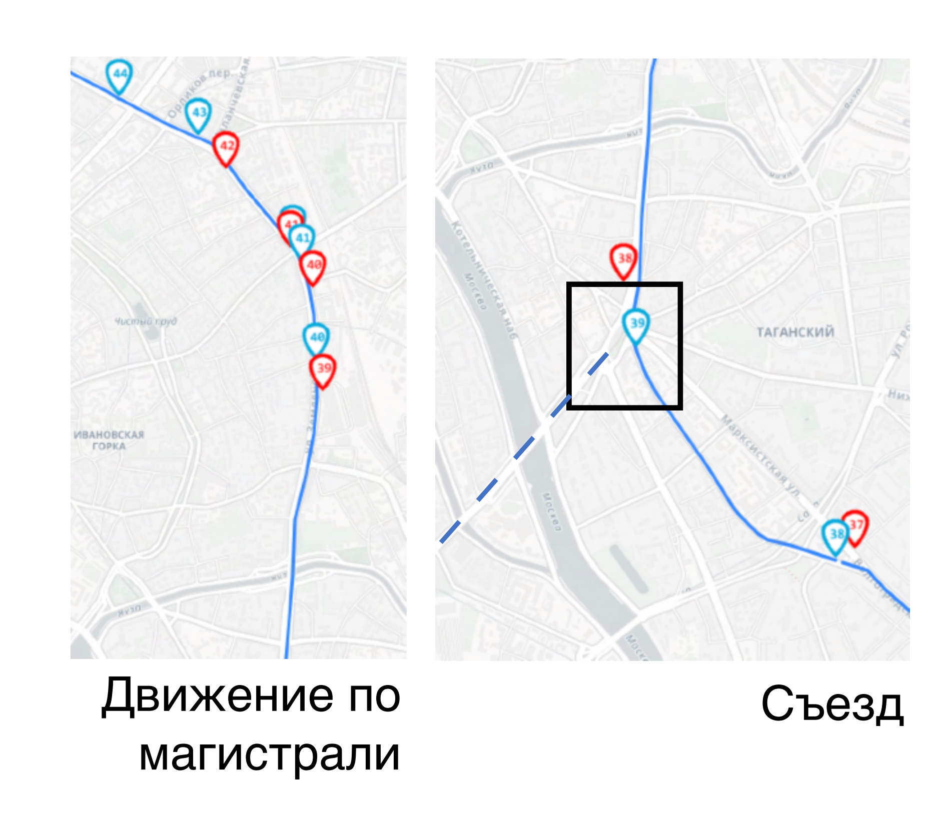
Съезд



Что происходит с перемещениями?

Рассмотрим подробнее перемещение на авто

- При перемещениях мы можем предугадывать точку назначения и прогнозировать путь
- За счет справочника соседства (фактического и эмпирического) можем предсказывать набор следующих БС (де-факто марковские цепи)
- Для каждого расположения существуют **степени свободы** — куда может перемещаться абонент



Движение по магистрали

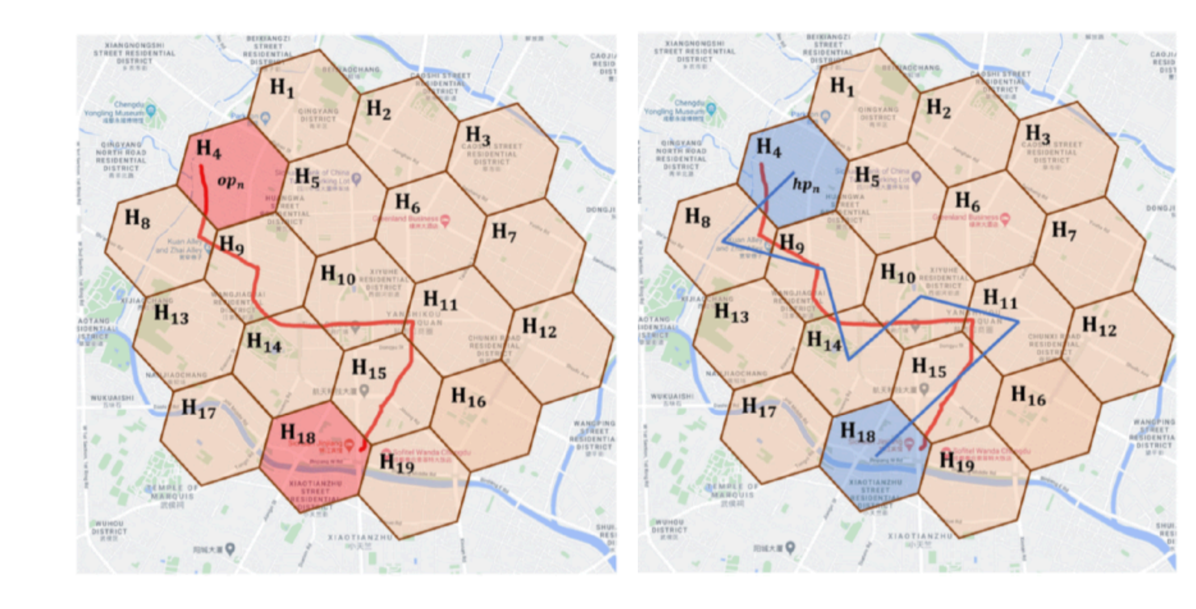
Съезд



Гексагональный дорожный граф

В состоянии движения нам не так важно, по какой именно дороге движется абонент

Граф тяжелый и ребра — ломанные => Упрощаем структуру и переходим на гексы





Гексагональный дорожный граф

В состоянии движения нам не так важно, по какой именно дороге движется абонент

Граф тяжелый и ребра — ломанные => Упрощаем структуру и переходим на гексы

1. Отсечем в подложке основные дороги



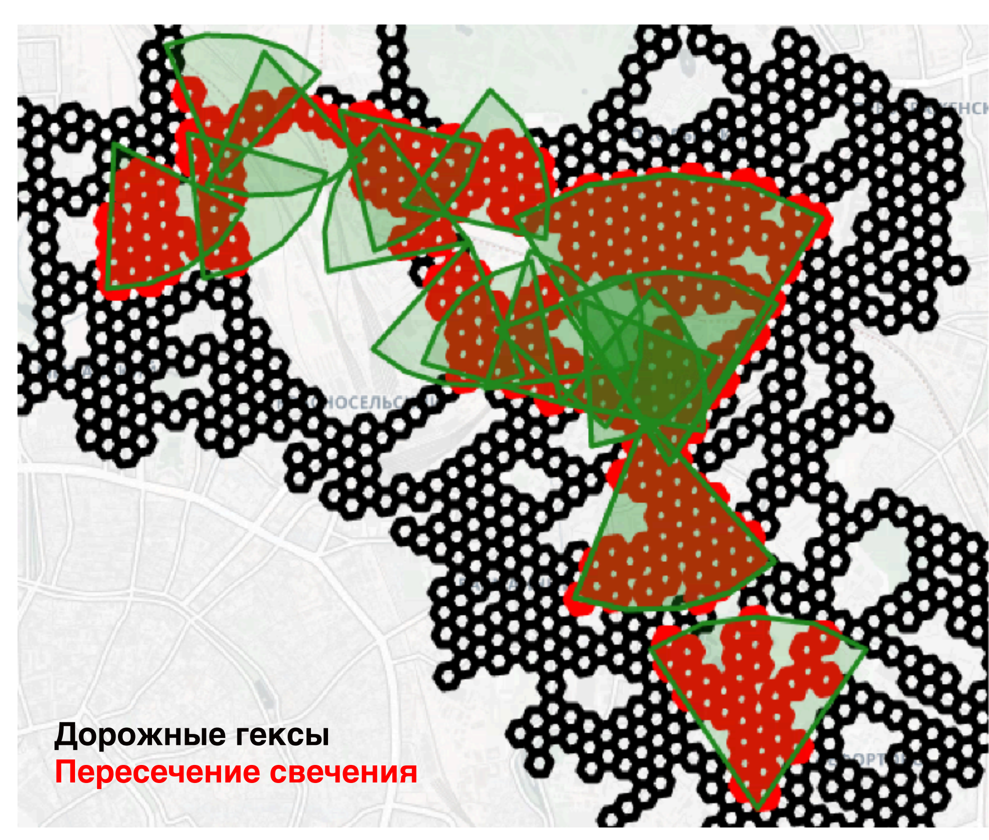


Гексагональный дорожный граф

В состоянии движения нам не так важно, по какой именно дороге движется абонент

Граф тяжелый и ребра — ломанные => Упрощаем структуру и переходим на гексы

1. Отсечем в подложке основные дороги
2. Для каждого гекса берем соседние и соединяем ребрами





Гексагональный дорожный граф

В состоянии движения нам не так важно, по какой именно дороге движется абонент

Граф тяжелый и ребра — ломанные => Упрощаем структуру и переходим на гексы

1. Отсечем в подложке основные дороги
2. Для каждого гекса берем соседние и соединяем ребрами
3. Получаем новый граф, где вершина — центроид гексагона

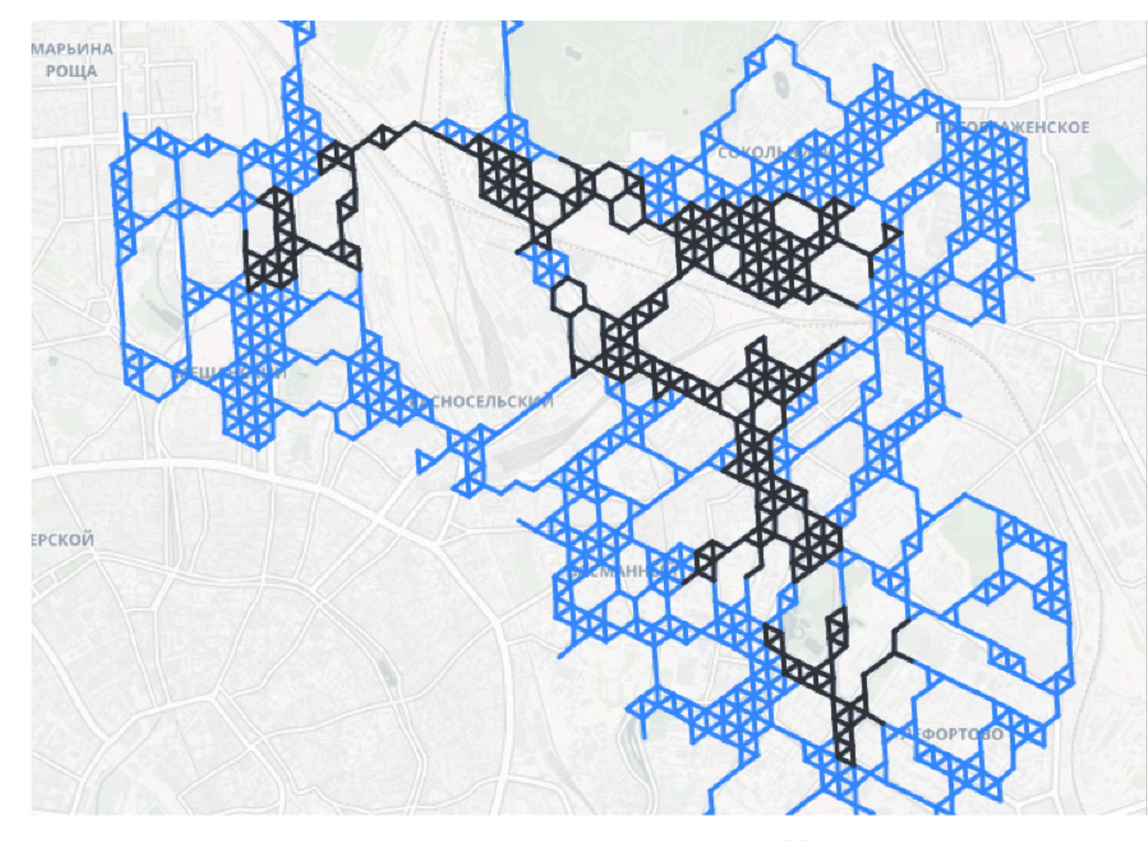




Схема построения гексагонального графа

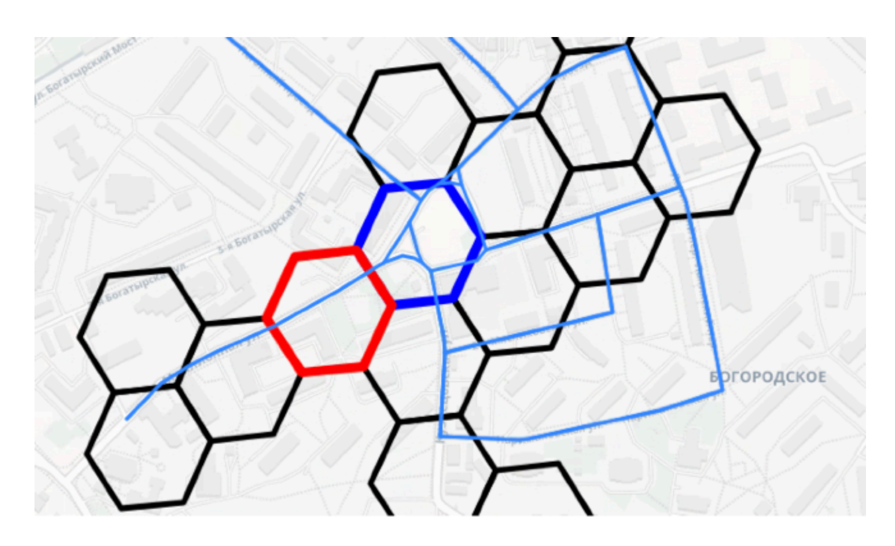




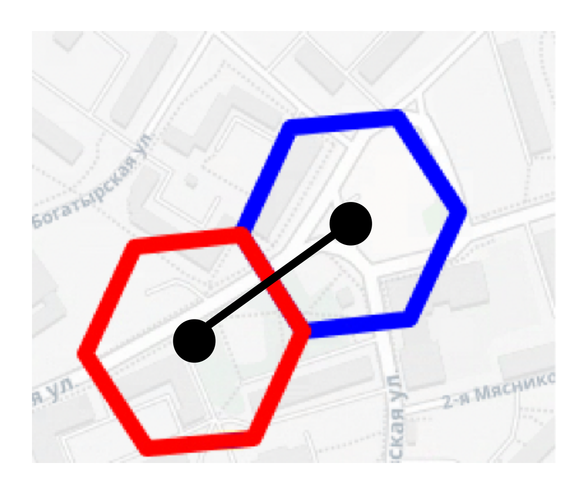
Пример дорожных гексов и дорог

Новые ребра





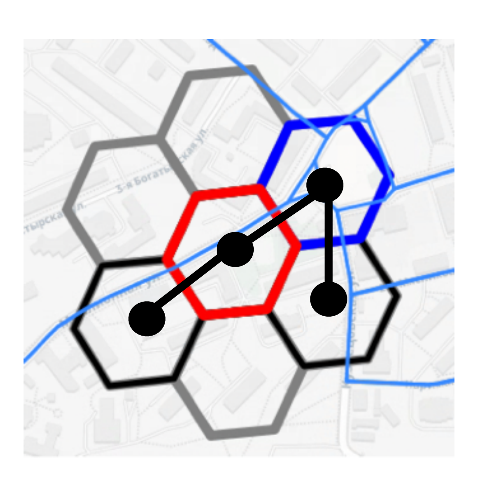
Пример дорожных гексов и дорог



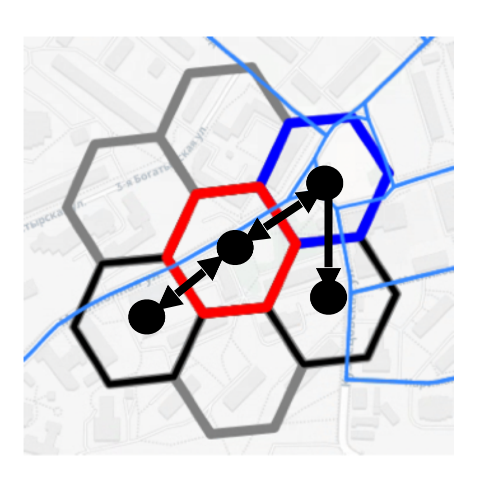
Новые ребра



Простейший граф с соединением соседних вершин



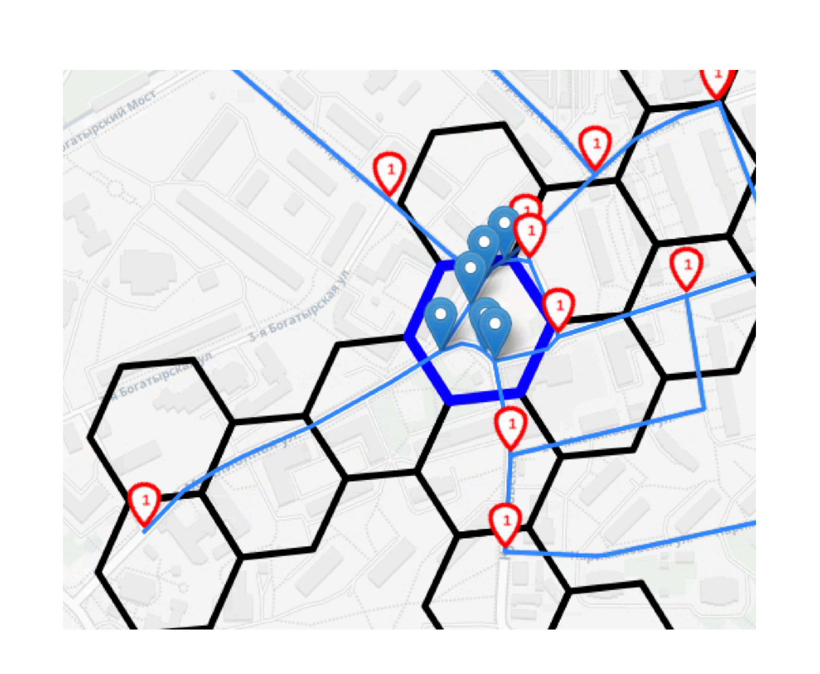
Учет дорог



Ориентированный взвешенный граф



Ориентированный гексагональный дорожный граф





Взяли соседние гексы и рассчитали среднее время / расстояние между комбинациями вершин



Создаем промежуточные вершины и считаем время



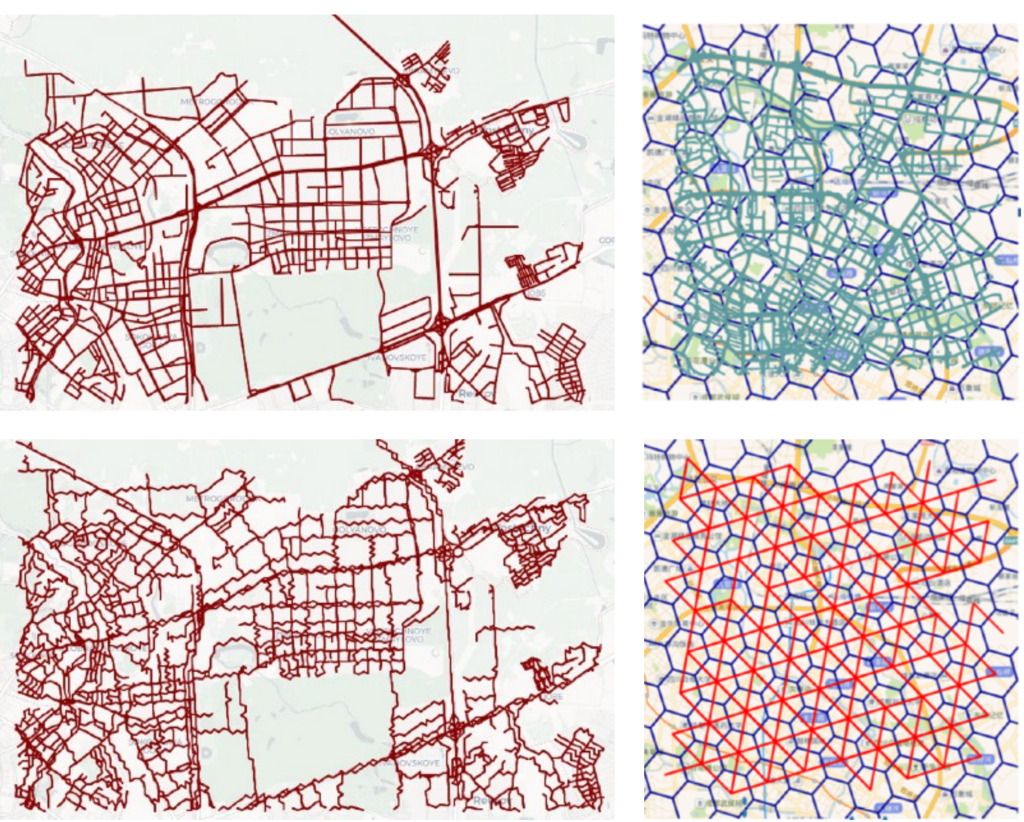
Проецируем на центроиды – новые вершины



Ориентированный гексагональный дорожный граф

1 Для каждой вершины графа в гексагоне смотрим соседние вершины

2 Для каждого гексагона попарно с соседним гексагоном берем подграф и считаем времена достижимости



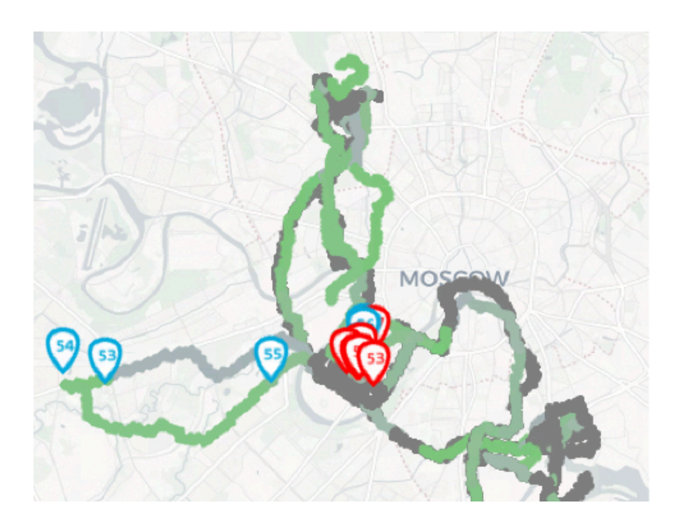


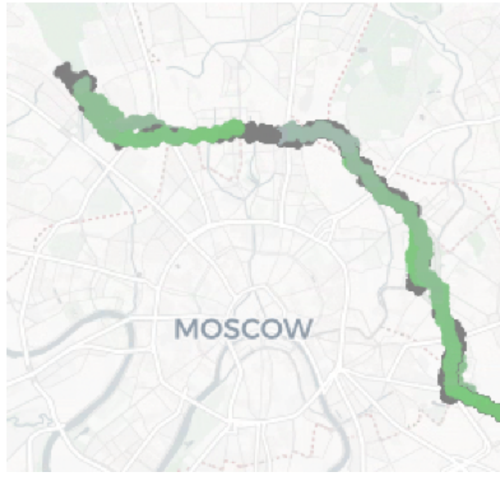
Как здесь считается ошибка

1. Считаем количество гексов между прогнозируемым и оригинальным маршрутом

2. Далее для заданного размера сетки считаем среднее расстояние между 2 гексами

3. Умножаем их число на это значение



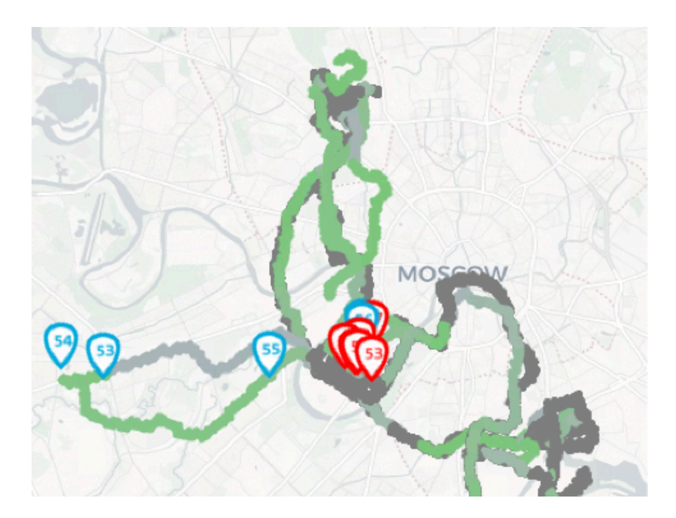


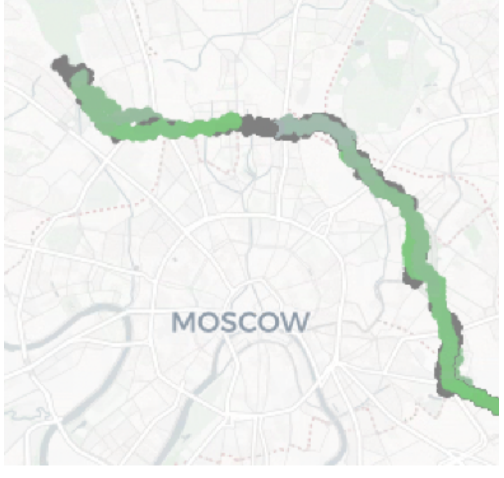




Как здесь считается ошибка

1. Считаем количество гексов между прогнозируемым и оригинальным маршрутом
2. Далее для заданного размера сетки считаем среднее расстояние между 2 гексами
3. Умножаем их число на это значение







Добились заявленного графового улучшения и получили быстрый рабочий метод



Next steps

1. Более детальный учет сегментации БС и точек притяжения

2. Учет временных факторов, пробок и скорости перемещения

3. Оптимизация расчетов маршрутов и использование предрассчитанных параметров



Что мы использовали в проекте

Spark

+ h3 libs
+ udf’s

Python

h3
networkx
osmnx
shapely
scipy.spatial

Visualizations

folium + osm



О чем сегодня поговорили?

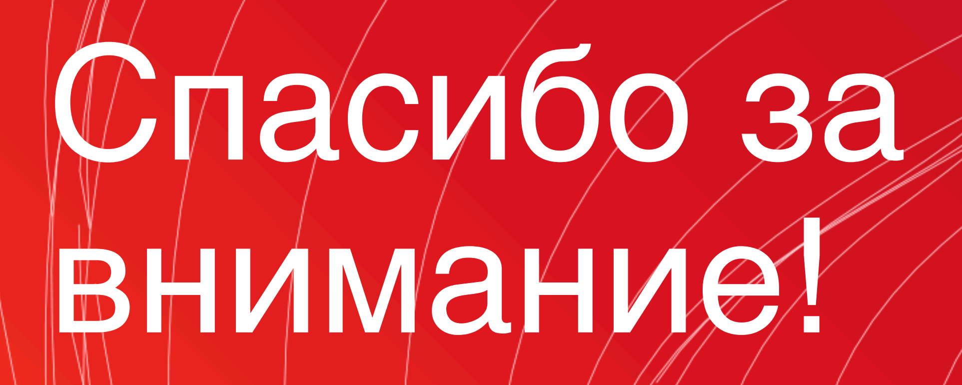
1 Как работать с нерегулярными событиями

2 Как удалять выбросы и выделять основную гео-информацию

3 Как строить маршруты по таким цепочкам

4 И как это делать быстрее и проще




Спасибо за
внимание!


HL
HighLoad++
2022

Контакты
Каледин Артем

Голосуйте за доклад



AVKaledin@beeline.ru

AVKaledin@mail.ru



ihitklif

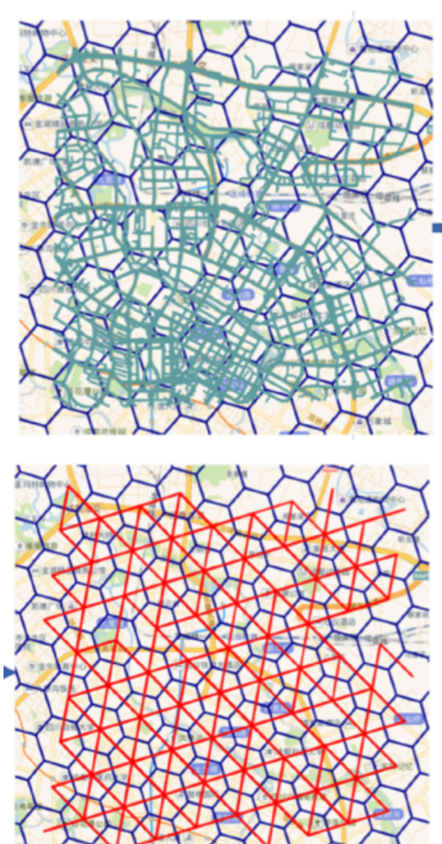


HighLoad++ 2022





Ориентированный гексагональный дорожный граф: описание шагов



1. Для каждой вершины графа в гексагоне смотрим соседние вершины
   A. Если вершина находится в соседнем гексе => OK
   B. Иначе => разбиваем ребро на линии и по ним создаем промежуточные точки в каждом гексагоне, чтобы построить гексагональные ребра
   C. Считаем время и расстояние, просто поделив на число гексагонов, либо производим расчет времени и расстояний по парам точек через маршрутизатор

2. Для каждого гексагона попарно с каждым соседним гексом возьмем подграф и рассчитаем по нему средние времена/расстояния достижимости и уже получим взвешенный гексагональный граф

